Features of a donor vector.

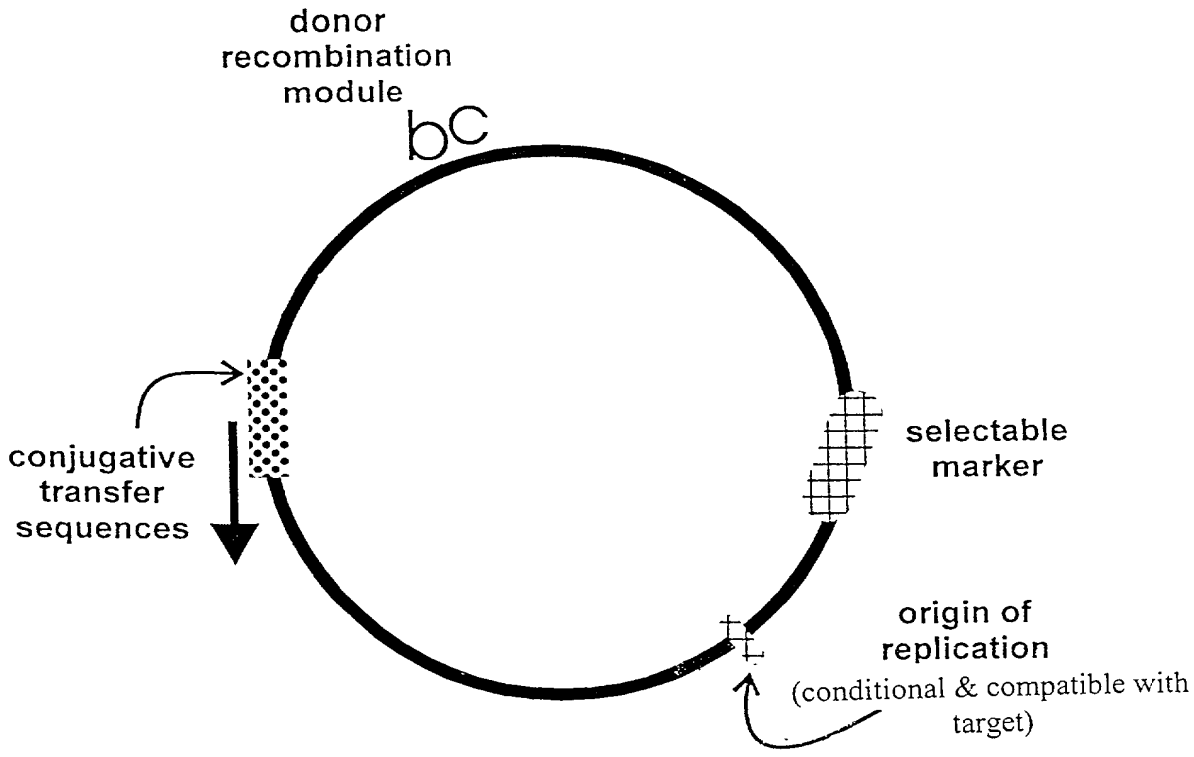
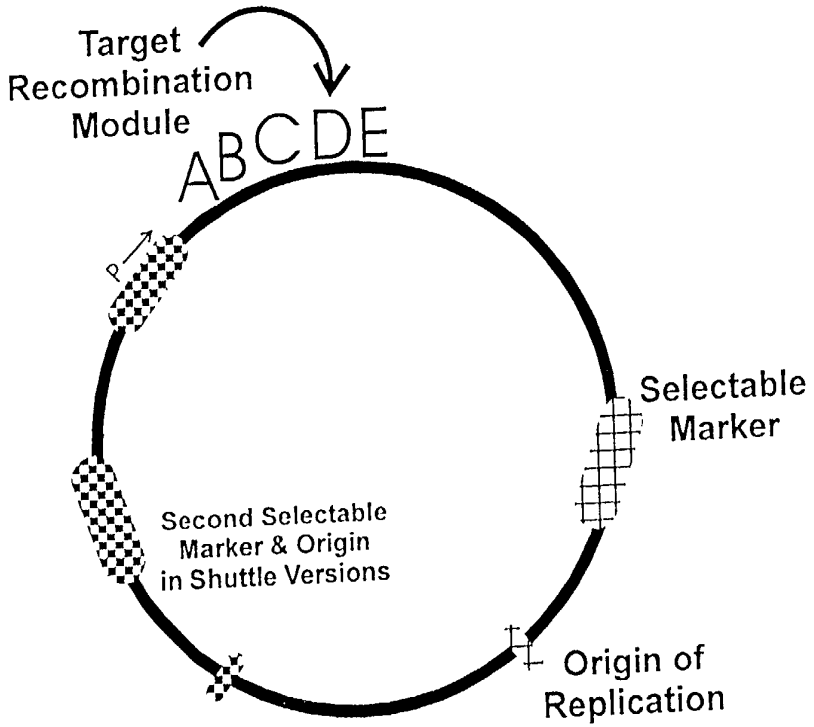


FIG. 1

Features of a target vector.



Additional Features of DGA

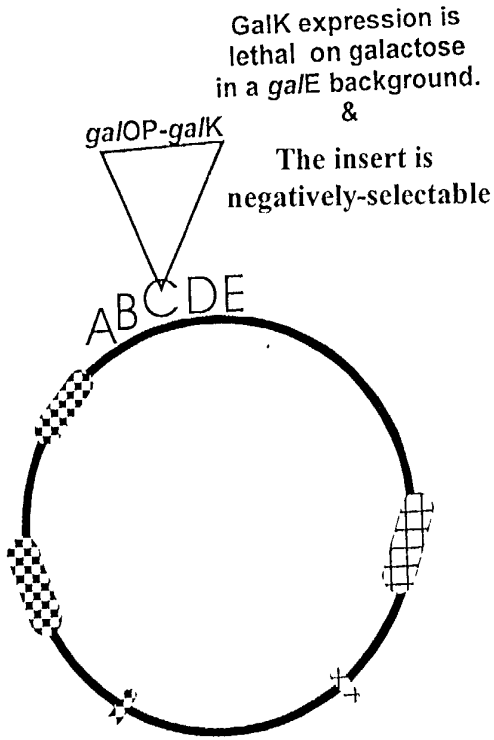


FIG. 2

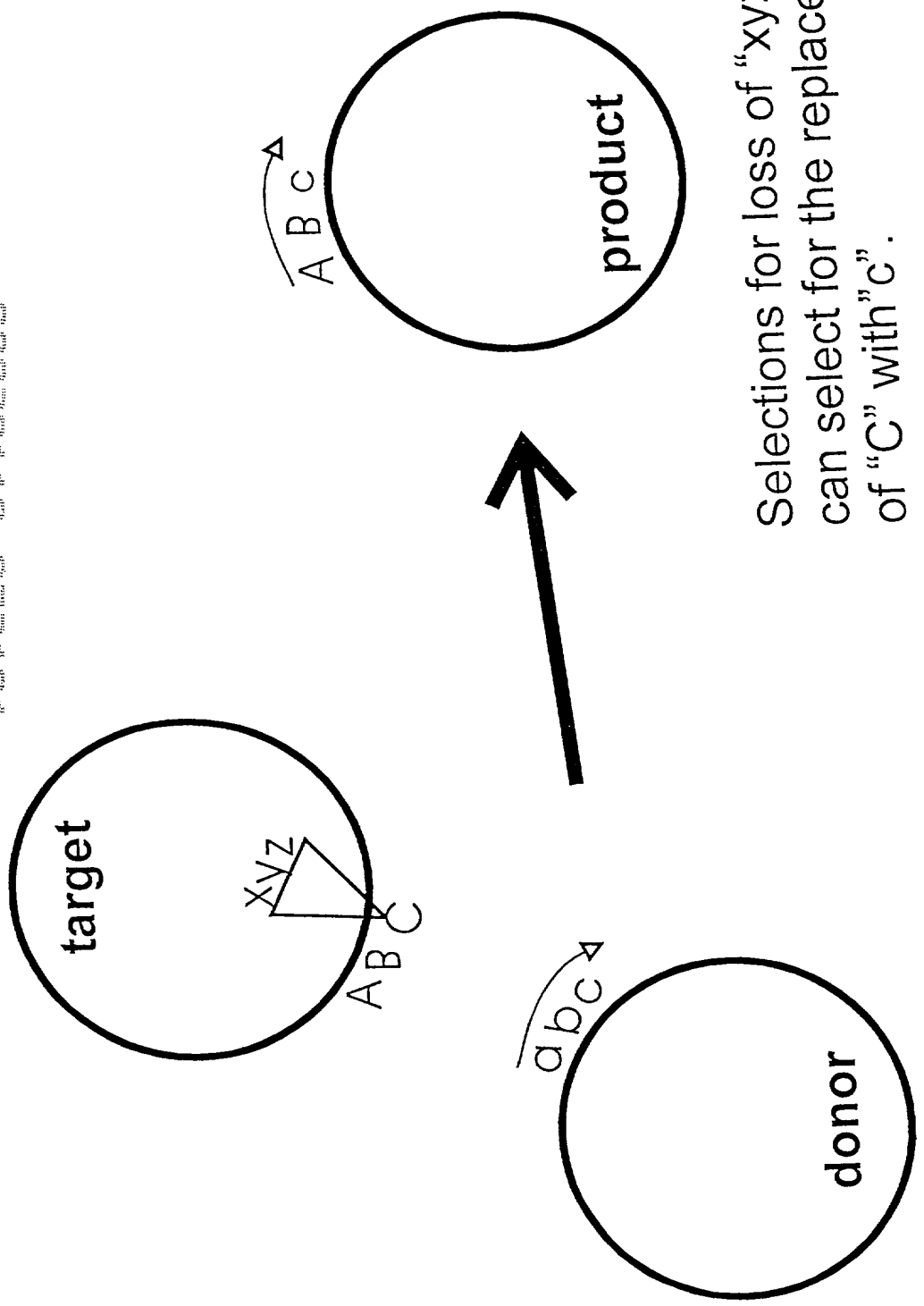
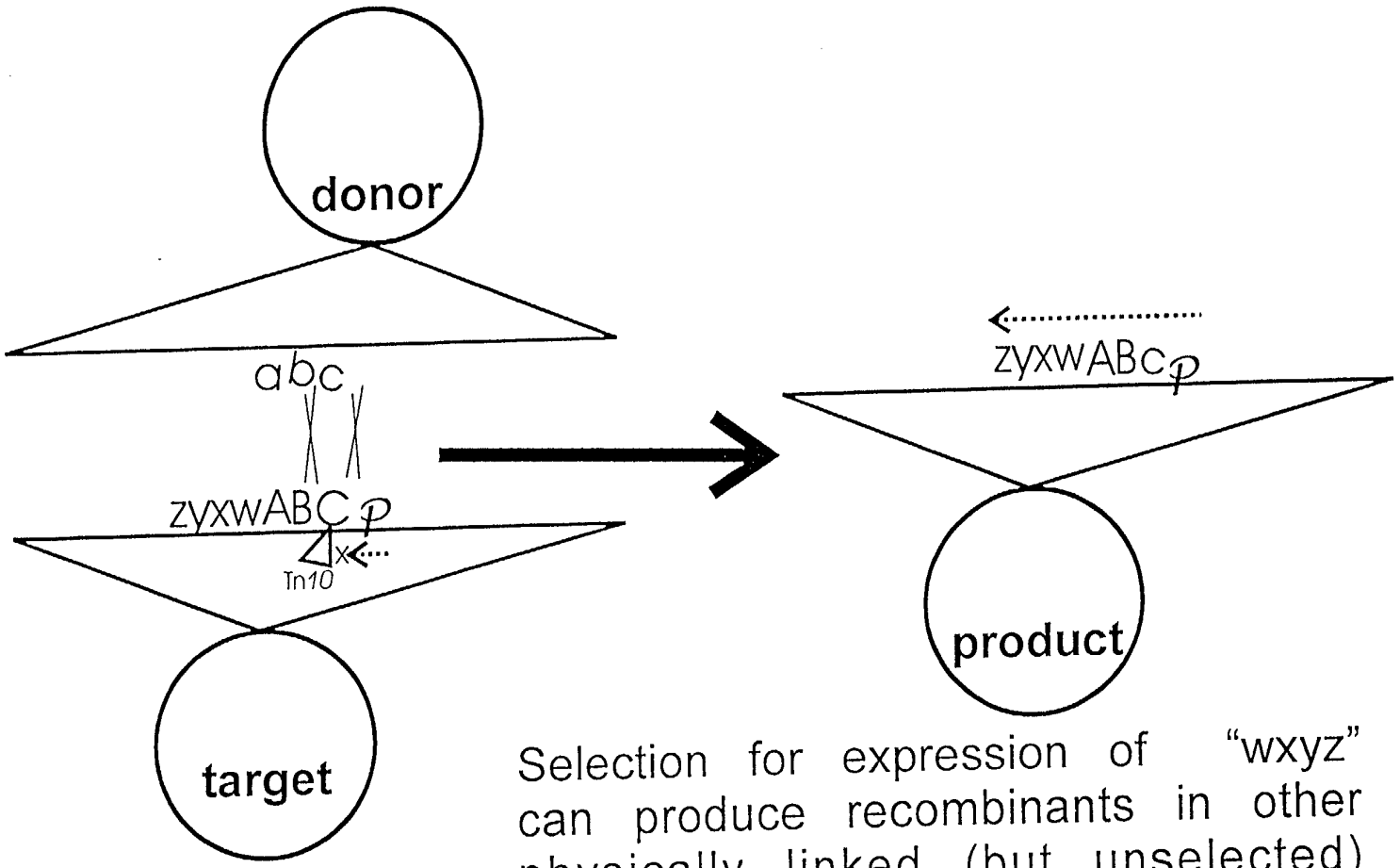
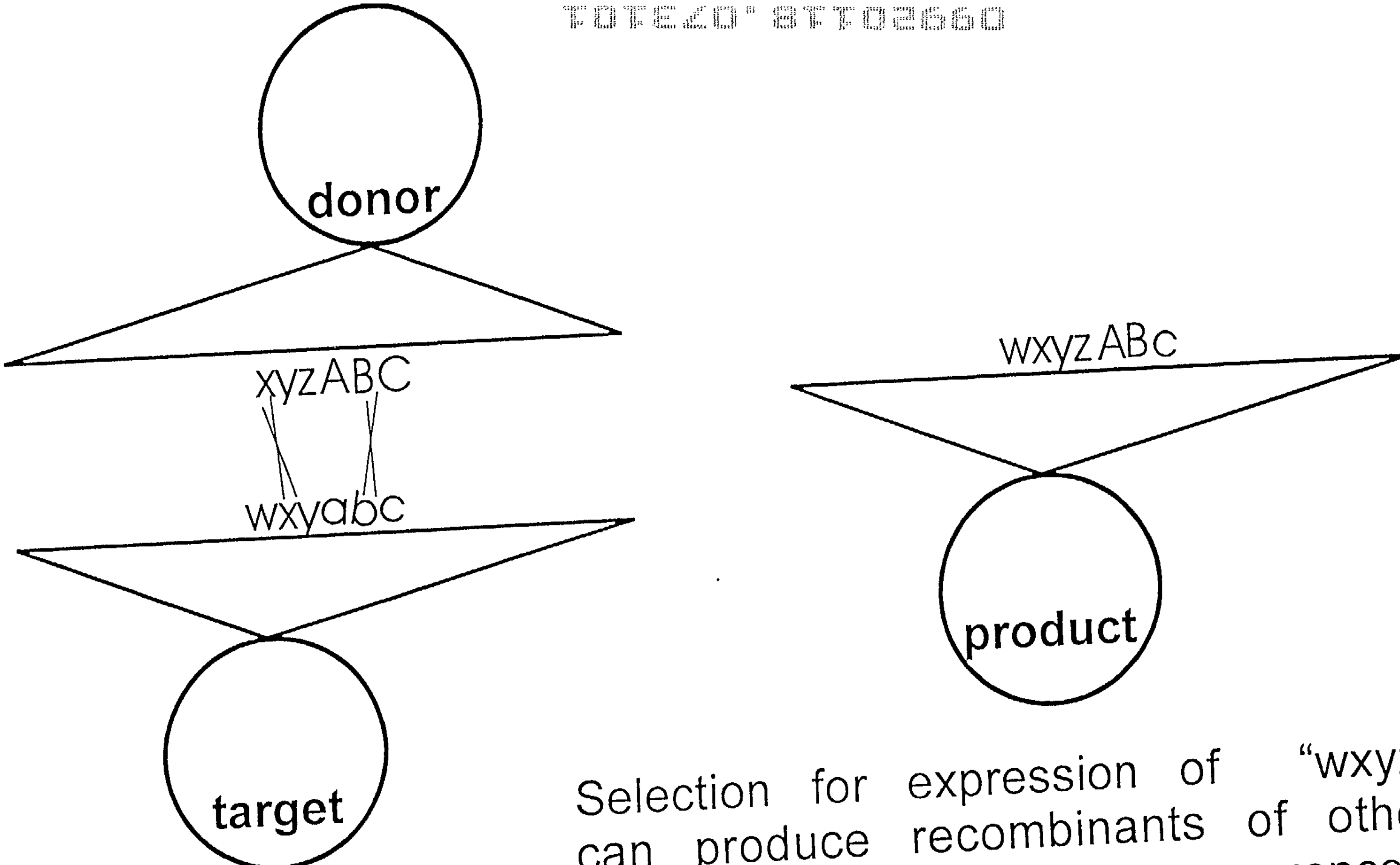

target
xyz
A B C
donor
a b c
product
A B c
Selections for loss of "xyz" can select for the replacement of "C" with "c".

FIG. 3



Selection for expression of "wxyz" can produce recombinants in other physically linked (but unselected) homologous sequences, such as, "ABc", in this example through removal of the polar Tn*10* sequences.

FIG. 4



Selection for expression of "wxyz" can produce recombinants of other physically linked homologous sequences, such as, "Abc" in the illustration.

FIG. 5

Directed Gene Assembly applied to mutagenesis

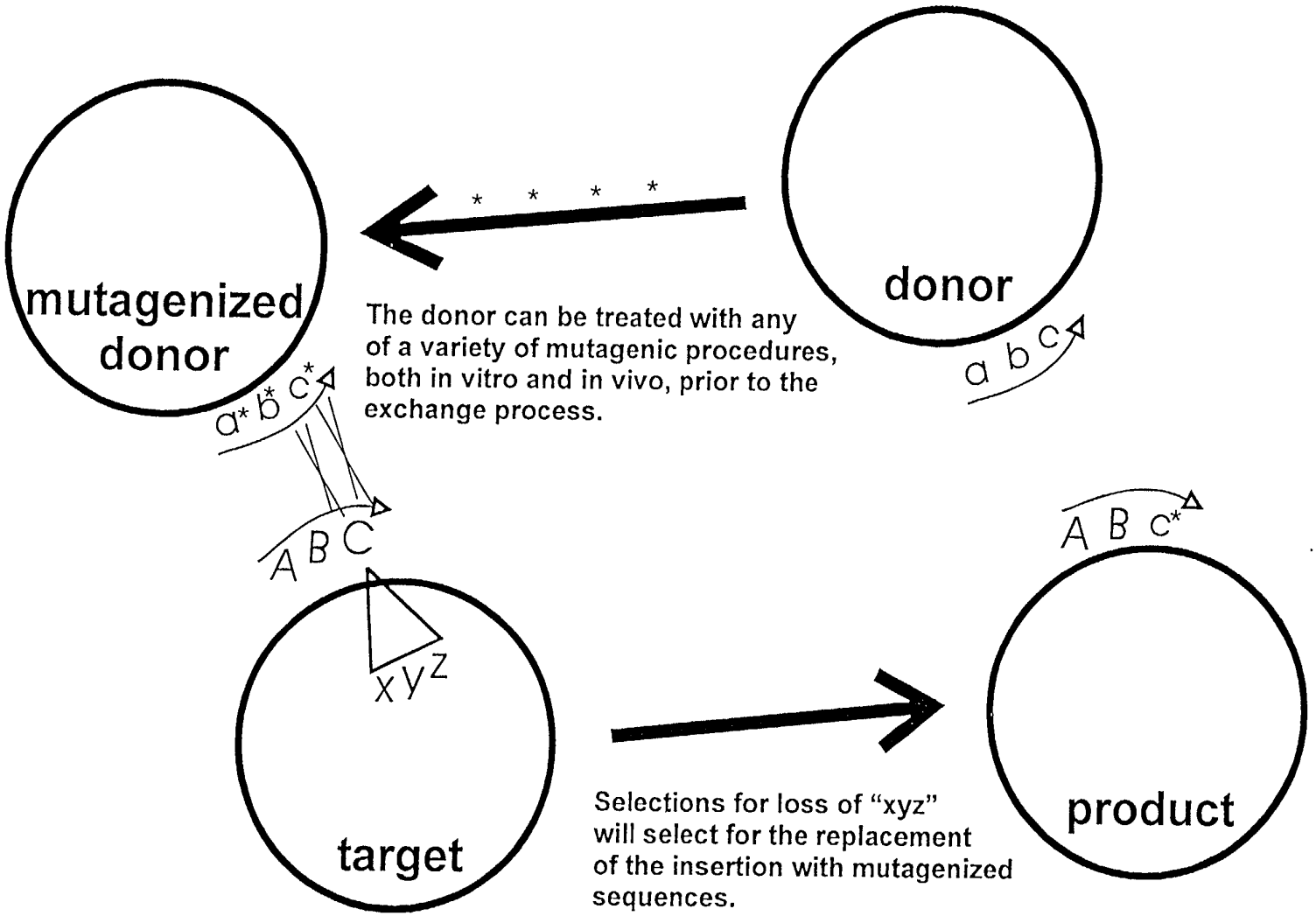
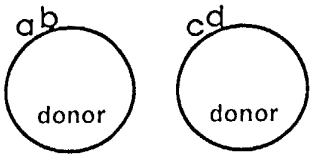


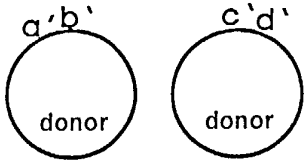
FIG. 6

Substrates

Target gene with family member # 1 (complete) with two recombination-targeting inserts in target vector.



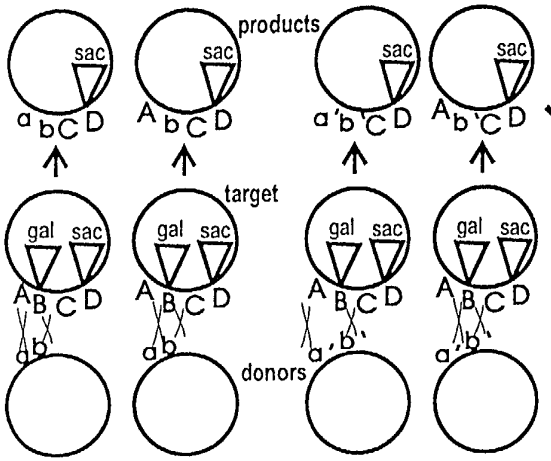
Fragments of family member #2 in donor vector library.



Fragments of family member #3 in donor vector library.

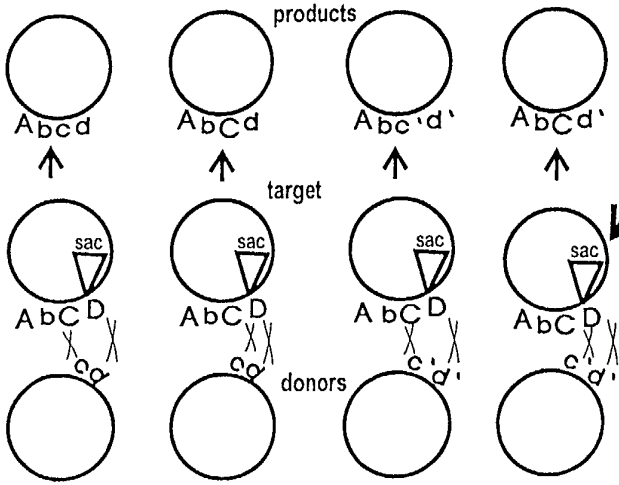
First Product Series
(galactose resistant)

Products of the first galactose resistant selected reaction still contains the sucrase insert disrupting the target.



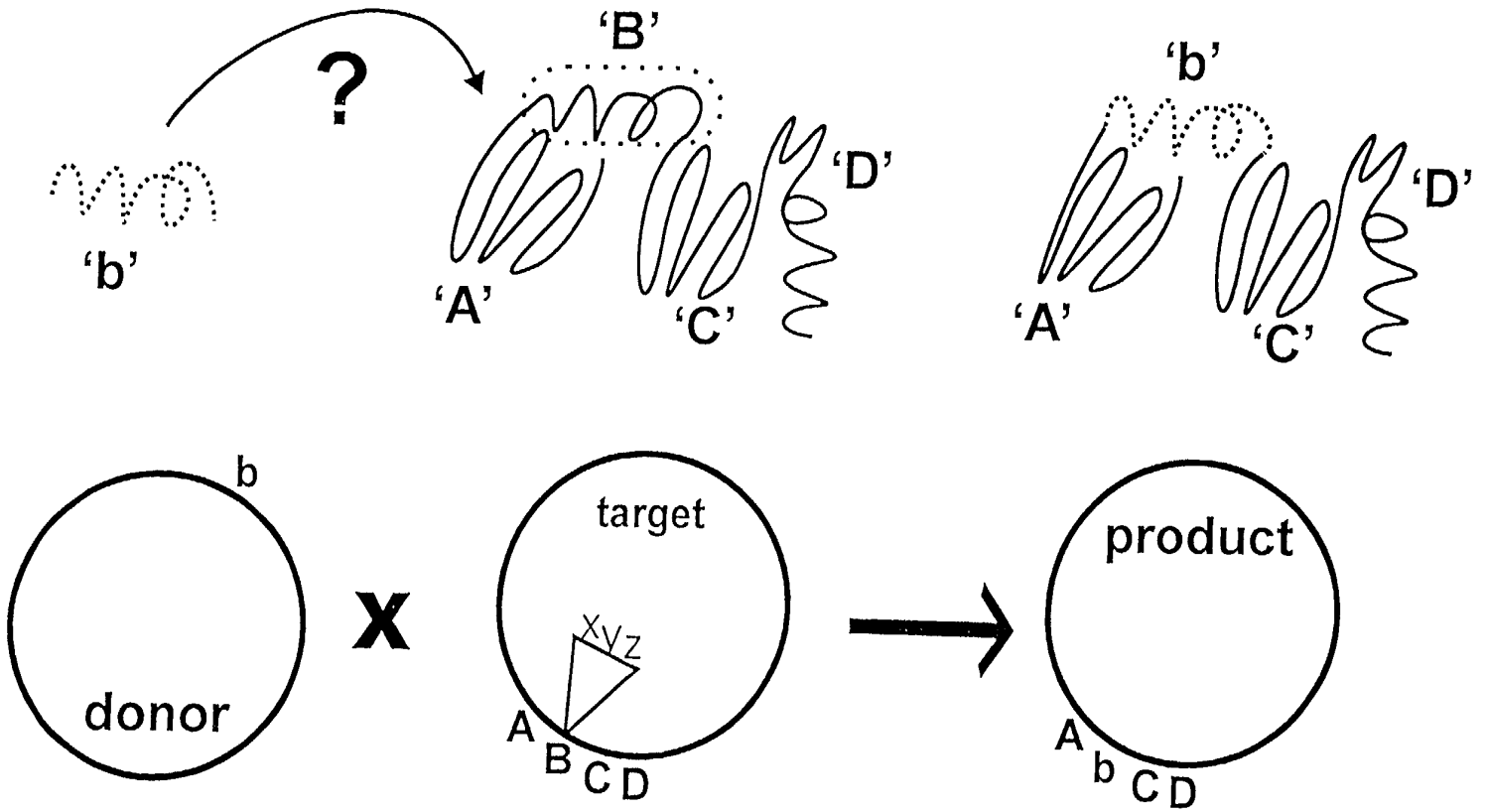
Second Product Series
(sucrose resistant)

After the second sucrose resistance selection intact genes are reconstructed.



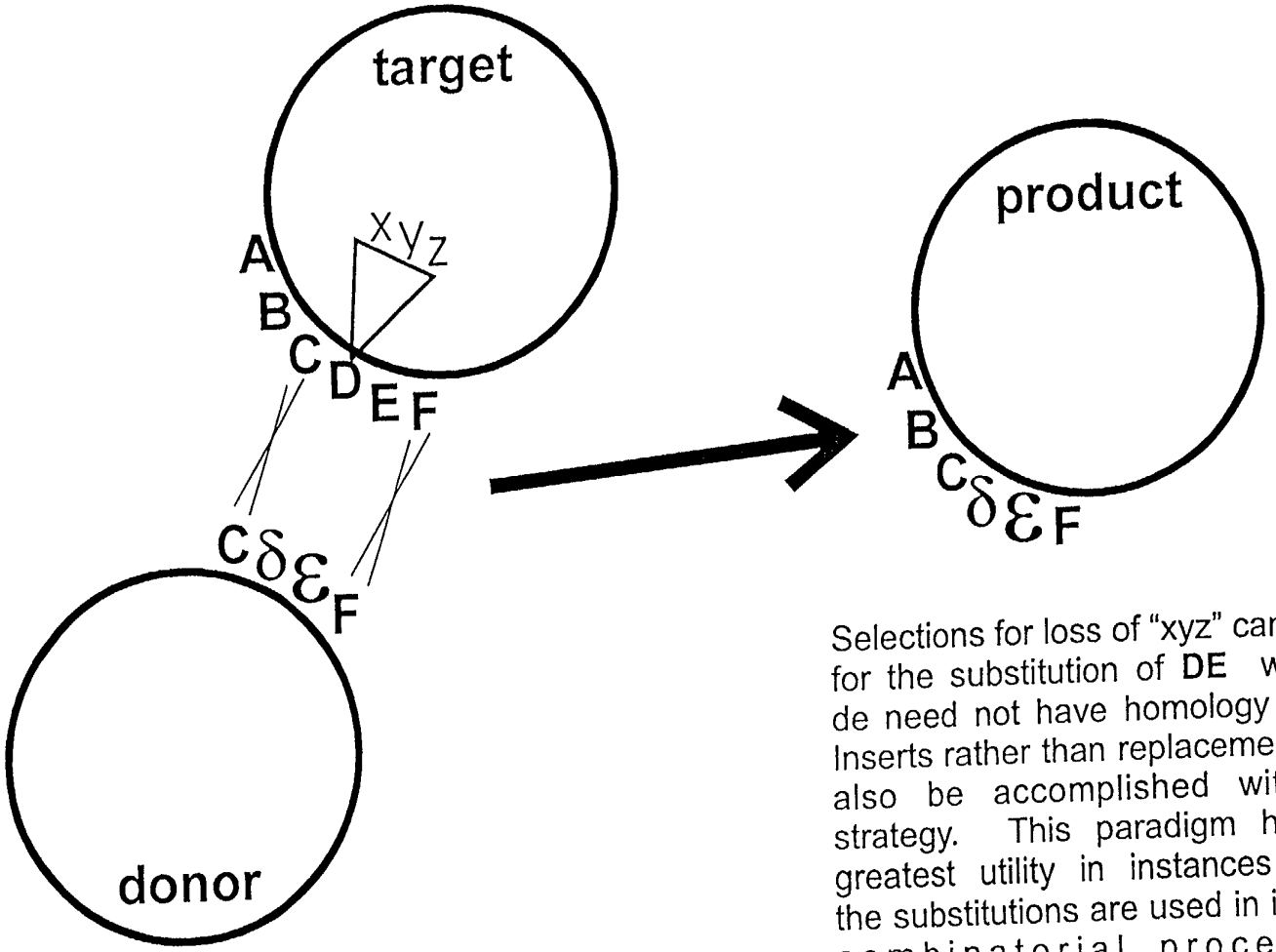
Products of the first reaction become the substrates of the second. Reactions with "AbCD" product are shown.

FIG. 7



An assay of product function will assess 'b's ability to substitute for 'B'. This paradigm executed with many target proteins and donor domains (and motifs) will identify a library of protein building blocks.

FIG. 8



Selections for loss of "xyz" can select for the substitution of **DE** with de. de need not have homology to DE. Inserts rather than replacements can also be accomplished with this strategy. This paradigm has the greatest utility in instances where the substitutions are used in iterative combinatorial processes.

FIG. 9

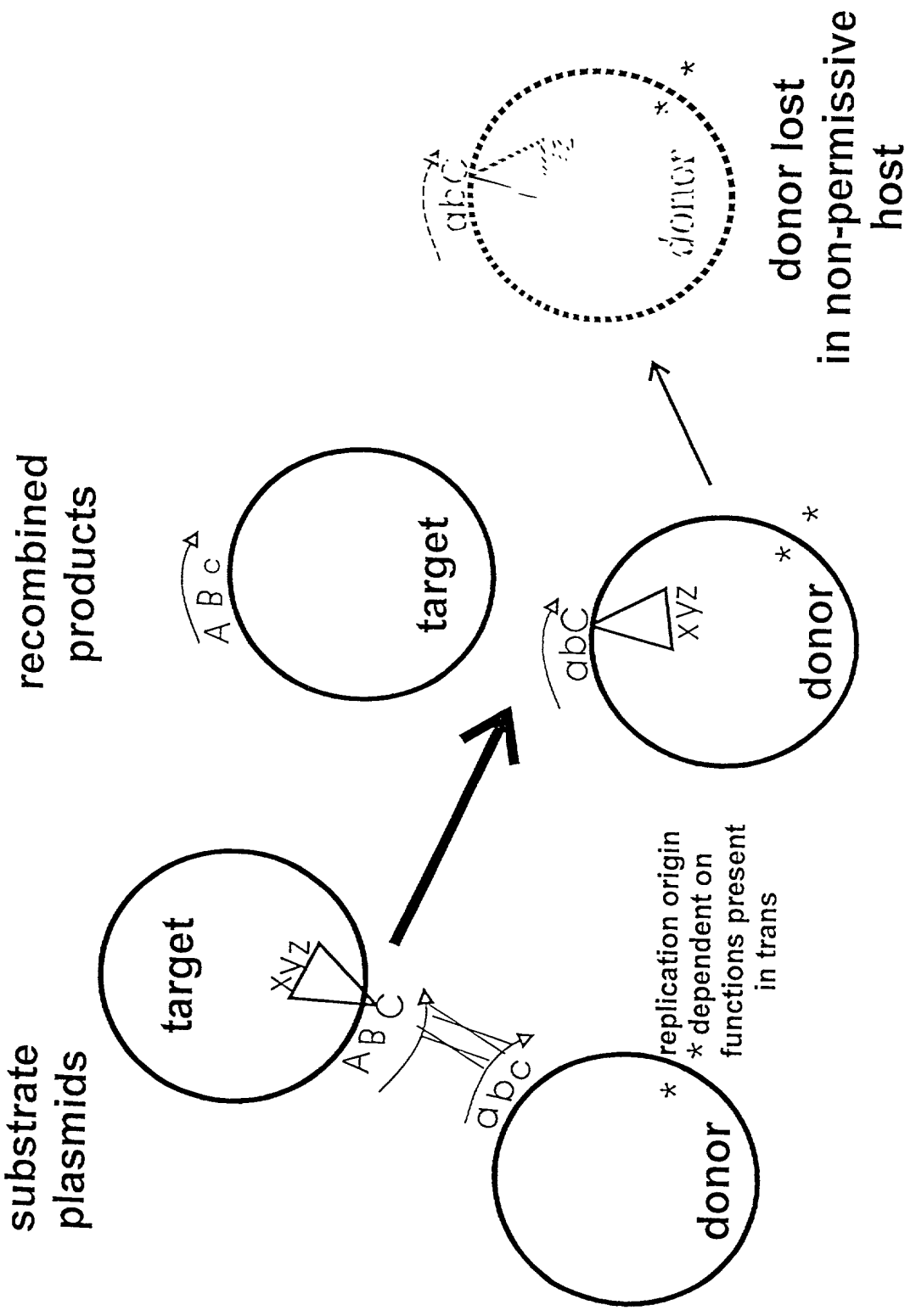

substrate plasmids
target
XYZ
ABC
abc
donor
* replication origin
* dependent on functions present in trans
recombined products
target
A B c
abC
XYZ
donor
* *
donor lost in non-permissive host
abC
donor
* *

FIG. 10

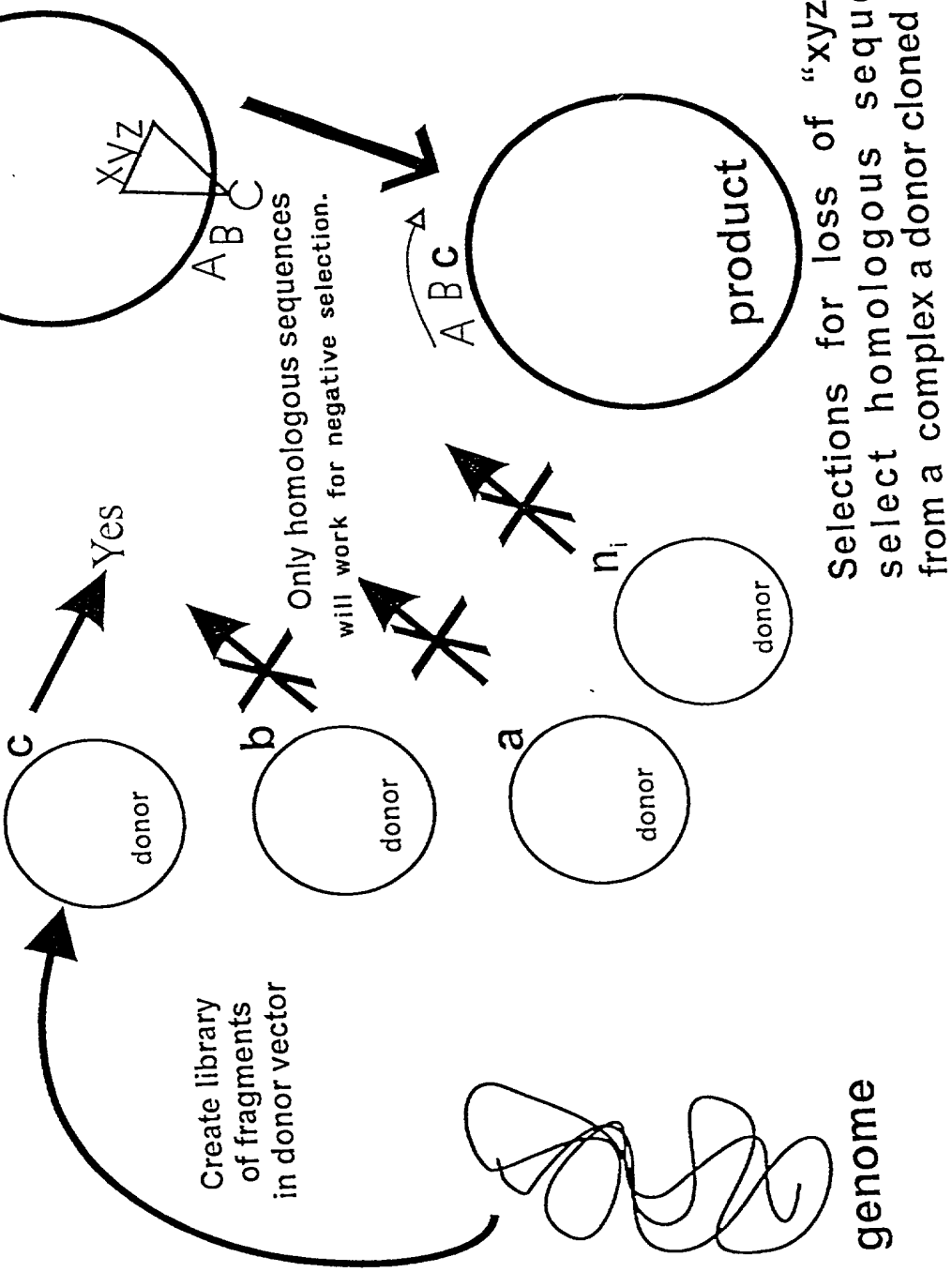

genome
Create library of fragments in donor vector
donor
donor
donor
donor
c
b
a
n_i
Yes
Only homologous sequences will work for negative selection.
target
XYZ
A B C
A B c
product
Selections for loss of "xyz" can select homologous sequences from a complex a donor cloned library.

FIG. 11

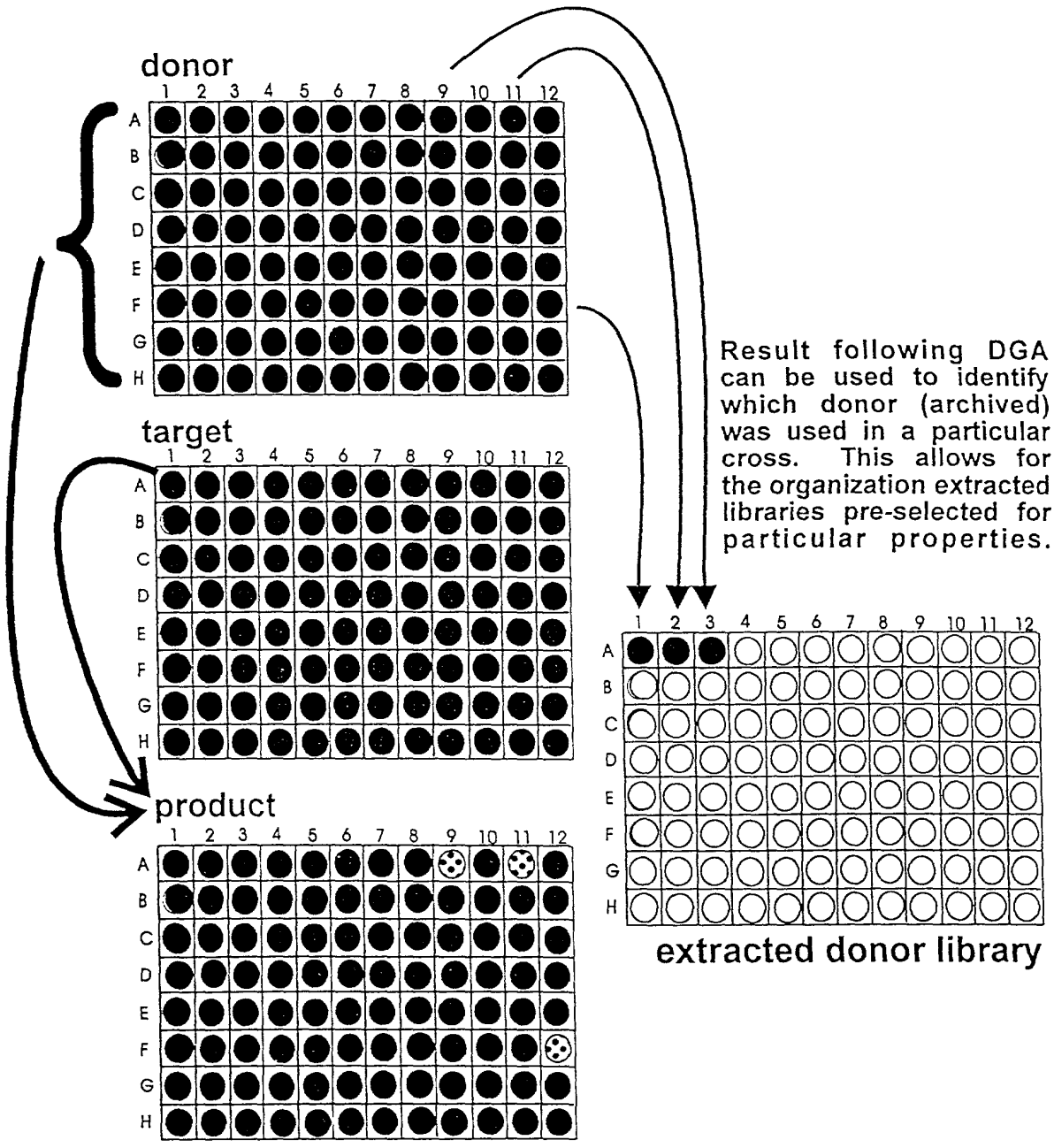

donor
target
product
Result following DGA can be used to identify which donor (archived) was used in a particular cross. This allows for the organization extracted libraries pre-selected for particular properties.
extracted donor library

FIG. 12

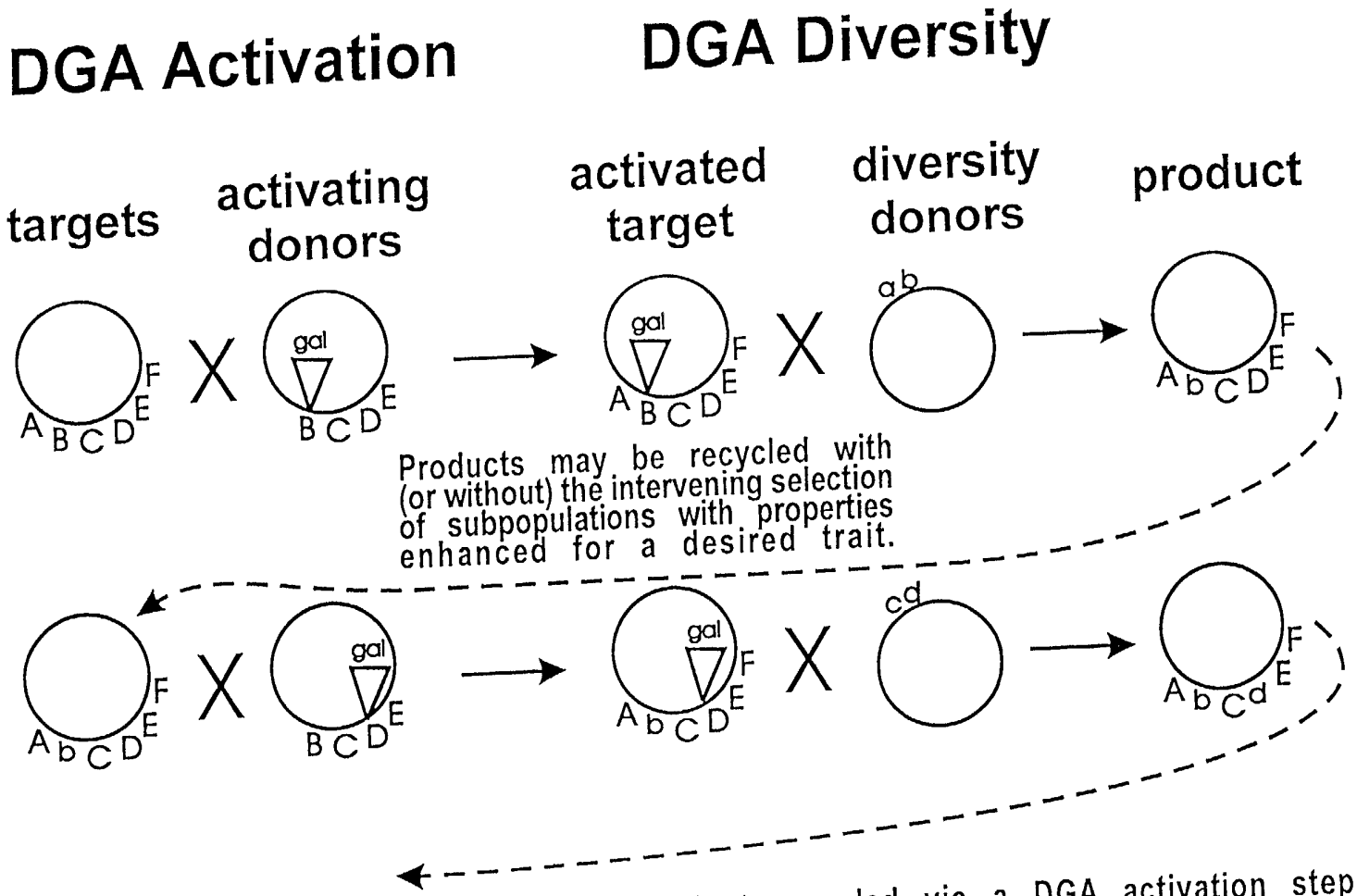


FIG. 13

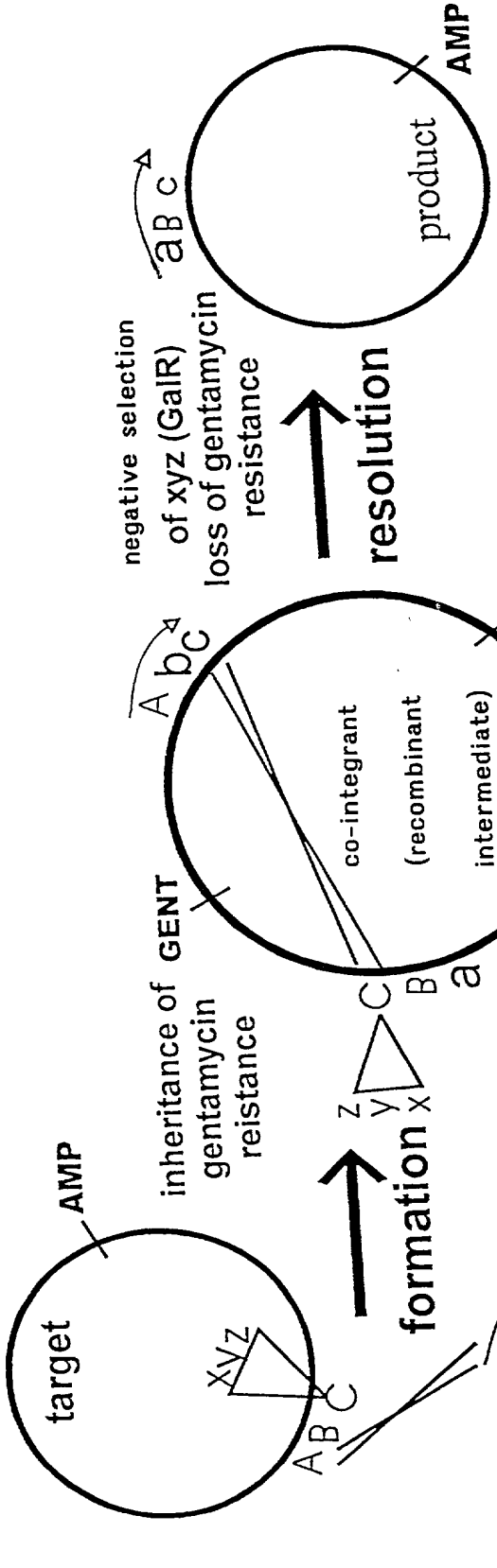

target
AMP
xyz
A B C
formation
inheritance of gentamycin reistance
z
y
x
C
B
a
GENT
co-integrant
(recombinant
intermediate)
A b C
AMP
negative selection
of xyz (GalR)
loss of gentamycin
resistance
resolution
a B c
product
AMP
a b C
donor
GENT

FIG. 14

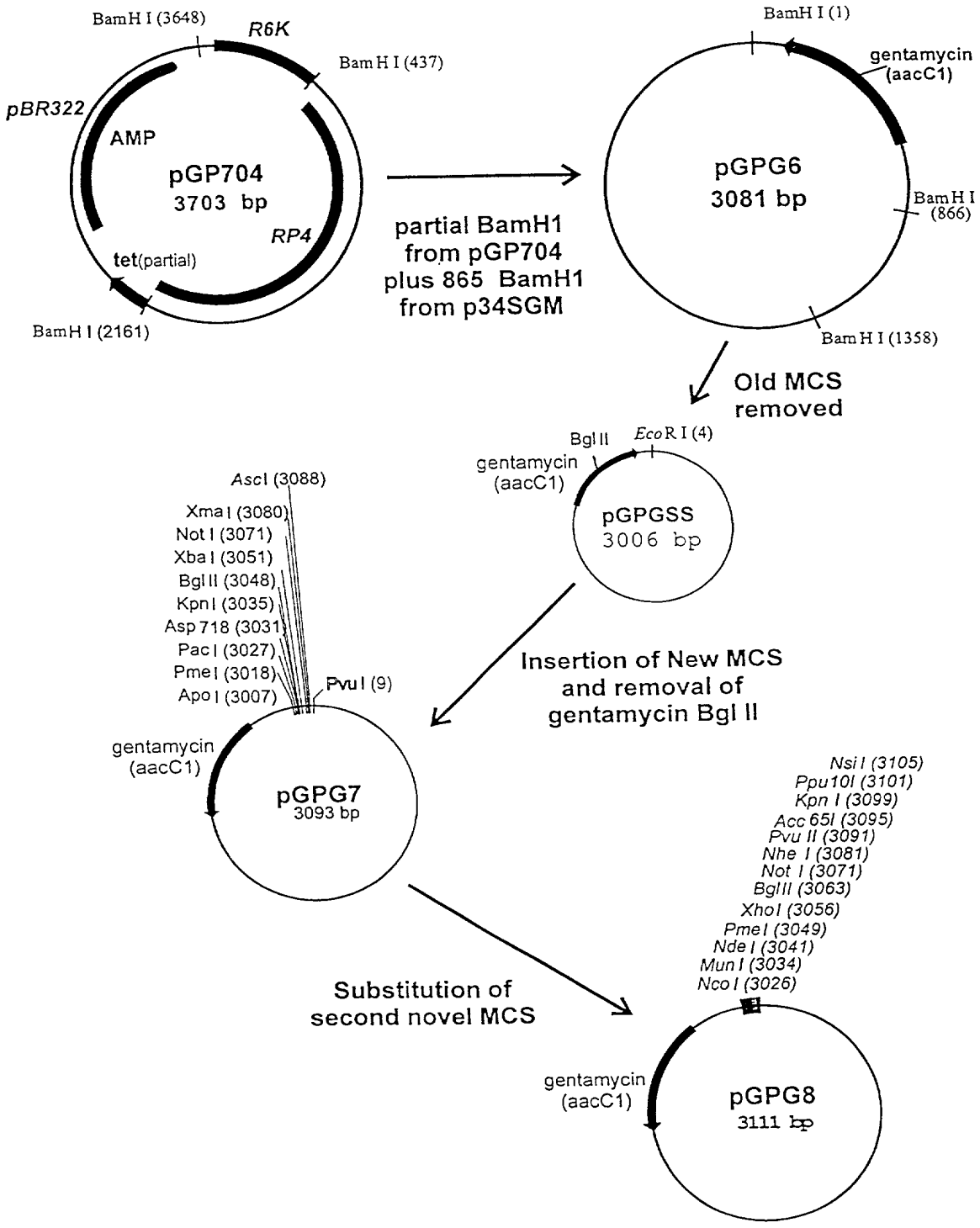

BamH I (3648)
R6K
BamH I (437)
pBR322
AMP
pGP704
3703 bp
RP4
tet(partial)
BamH I (2161)
partial BamH1
from pGP704
plus 865 BamH1
from p34SGM
BamH I (1)
gentamycin
(aacC1)
pGPG6
3081 bp
BamH I
(866)
BamH I (1358)
Old MCS
removed
BglII
EcoR I (4)
gentamycin
(aacC1)
pGPGSS
3006 bp
AscI (3088)
Xma I (3080)
Not I (3071)
Xba I (3051)
BglII (3048)
Kpn I (3035)
Asp 718 (3031)
Pac I (3027)
Pme I (3018)
Apo I (3007)
PvuI (9)
Insertion of New MCS
and removal of
gentamycin Bgl II
gentamycin
(aacC1)
pGPG7
3093 bp
Nsi I (3105)
Ppu10I (3101)
Kpn I (3099)
Acc 65I (3095)
Pvu II (3091)
Nhe I (3081)
Not I (3071)
BglII (3063)
XhoI (3056)
Pme I (3049)
Nde I (3041)
Mun I (3034)
Nco I (3026)
Substitution of
second novel MCS
gentamycin
(aacC1)
pGPG8
3111 bp

FIG. 15

```
 +1  I  K  A  D  K  V  Q  A  Q  G  F  K  G  A  N  V  K  V  A  V  L
  1  ATTAAAGCG GACAAAGTG CAGGCTCAA GGCTTTAAG GGAGCGAAT GTAAAAGTA GCCGTCCTG

 +1  D  T  G  I  Q  A  S  H  P  D  L  N  V  V  G  G  A  S  F  V  A
 64  GATACAGGA ATCCAAGCT TCTCATCCG GACTTGAAC GTAGTCGGC GGAGCAAGC TTTGTGGCT

 +1  G  E  A  Y  N  T  D  G  N  G  H  G  A  H  V  A  G  T  V  A  A
127  GGCGAAGCT TATAACACC GACGGCAAC GGACACGGC GCACATGTT GCCGGTACA GTAGCTGCG

 +1  L  D  N  T  T  G  V  L  G  V  A  P  S  V  S  L  Y  A  V  K  V
190  CTTGACAAT ACAACGGGT GTATTAGGC GTTGCGCCA AGCGTATCC TTGTACGCG GTTAAAGTA

 +1  L  N  S  S  G  S  G  S  Y  S  G  I  V  S  G  I  E  W  A  T  T
253  CTGAATTCA AGCGGAAGC GGATCATAC AGCGGCATT GTAAGCGGA ATCGAGTGG GCGACAACA

 +1  N  G  M  D  V  I  N  M  S  L  G  G  A  S  G  S  T  A  M  K  Q
316  AACGGCATG GATGTTATC AATATGAGC CTTGGGGGA GCATCAGGC TCGACAGCG ATGAAACAG

 +1  A  V  D  N  A  Y  A  K  G  V  V  V  V  A  A  A  G  N  S  G  S
379  GCAGTCGAC AATGCATAT GCAAAAGGG GTTGTCGTT GTAGCTGCA GCAGGGAAC AGCGGATCT
```

DNA Sequence 5A20 *B.licheniformis* isolate

```
 +1  I  K  A  P  A  L  H  S  Q  G  Y  T  G  S  N  V  K  V  A  V  I
  1  ATTAAAGCG CCGGCTCTT CACTCTCAA GGCTACACA GGTTCTAAC GTAAAAGTA GCCGTAATT

 +1  D  S  G  I  D  S  S  H  P  D  L  N  V  R  G  G  A  S  F  V  P
 64  GACAGCGGA ATTGACTCT TCTCATCCT GACTTGAAC GTCAGAGGC GGAGCAAGC TTCGTACCT

 +1  S  E  T  N  P  Y  Q  D  G  S  S  H  G  T  H  V  A  G  T  V  A
127  TCTGAAACA AACCCATAC CAAGATGGC AGTTCTCAC GGCACACAT GTAGCCGGT ACGGTTGCC

 +1  A  L  N  N  S  I  G  V  L  G  V  A  P  N  A  S  L  Y  A  V  K
190  GCACTTAAT AACTCAATC GGTGTTTTG GGCGTAGCG CCAAACGCA TCGTTATAT GCAGTAAAA

 +1  V  L  D  S  T  G  N  G  Q  Y  S  W  I  I  N  G  I  E  W  A  I
253  GTTCTTGAT TCAACAGGA AACGGCCAA TACAGCTGG ATTATTAAC GGCATTGAG TGGGCCATT

 +1  S  N  K  M  D  V  I  N  M  S  L  G  G  P  S  G  S  T  A  L  K
316  TCCAACAAA ATGGACGTG ATTAACATG AGCCTTGGC GGACCTTCT GGTTCTACA GCTTTGAAA
 +1  S  V  V  D  R  A  V  A  S  G  I  V  V  V  A  A  A  G  N  E  G
379  TCAGTCGTT GATAGAGCC GTAGCCAGC GGTATCGTC GTTGTTGCT GCAGCCGGA AATGAAGGC

 +1  T  S  G  S  S  S  T  I  G  Y  P  A  K  Y  P  S  T  I  A  V  G
442  ACTTCCGGA AGCTCAAGC ACAATCGGC TATCCTGCA AAATATCCT TCTACCATT GCGGTAGGT

 +1  A  V  N  S  S  N  Q  R  G  S  F  S  S  V  G  P  E  L  E  V  M
505  GCGGTAAAC AGCAGCAAC CAAAGAGGT TCATTCTCA AGCGTAGGT CCTGAGCTT GAAGTCATG

 +1  A  P  G
568  GCTCCTGGC
```

DNA Sequence 3A13 *B.subtilis* isolate

FIG. 16

```
      M  M  R   K  K  S   F  W  L   G  M  L   T  A  L   M  L  V   F  T  M
1     ATGATGAGG AAAAAGAGT TTTTGGCTT GGGATGCTG ACGGCCTTA ATGCTCGTG TTCACGATG

      A  F  S   D  S  A   S  A  A   Q  P  A   K  N  V   E  K  D   Y  I  V
64    GCCTTCAGC GATTCCGCG TCTGCTGCT CAGCCGGCG AAAAATGTT GAAAAGGAT TATATTGTC

      G  F  K   S  G  V   K  T  A   S  V  K   K  D  I   I  K  E   S  G  G
127   GGATTTAAG TCGGGAGTG AAAACCGCA TCCGTCAAA AAGGACATC ATCAAAGAG AGCGGCGGA

      K  V  D   K  Q  F   R  I  I   N  A  A   K  A  K   L  D  K   E  A  L
190   AAAGTGGAC AAGCAGTTT AGAATCATC AACGCGGCA AAAGCGAAG CTAGACAAA GAAGCGCTT

      E  E  V   K  N  D   P  D  V   A  Y  V   E  E  D   H  V  A   H  A  L
253   GAGGAAGTC AAAAATGAT CCGGATGTC GCTTATGTG GAAGAGGAT CACGTAGCT CATGCTTTG

      A  Q  T   V  P  Y   G  I  P   L  I  K   A  D  K   V  Q  A   Q  G  Y
316   GCGCAAACC GTTCCTTAC GGCATTCCT CTCATTAAA GCGGACAAA GTGCAGGCT CAAGGCTAC

      K  G  A   N  V  K   V  A  V   L  D  T   G  I  Q   A  S  H   P  D  L
379   AAGGGAGCG AACGTAAAA GTCGCCGTC CTGGATACA GGAATCCAA GCTTCTCAT CCGGACTTG

      N  V  V   G  G  A   S  F  V   A  G  E   A  Y  N   T  D  G   N  G  H
442   AACGTAGTC GGCGGAGCA AGCTTCGTA GCTGGCGAA GCTTATAAC ACCGACGGC AACGGACAC

      G  T  H   V  A  G   T  V  A   A  L  D   N  T  T   G  V  L   G  V  A
505   GGCACGCAT GTTGCCGGT ACAGTAGCT GCGCTTGAC AATACAACG GGTGTATTA GGCGTTGCG

      P  N  V   S  L  Y   A  V  K   V  L  N   S  S  G   S  G  S   Y  S  G
568   CCGAACGTA TCCTTGTAC GCGGTTAAA GTGCTGAAT TCAAGCGGA AGCGGATCT TACAGCGGC

      I  V  S   G  I  E   W  A  T   T  N  G   M  D  V   I  N  M   S  L  G
631   ATTGTAAGC GGAATCGAG TGGGCGACG ACAAACGGC ATGGATGTT ATCAACATG AGCCTTGGA

      G  P  S   G  S  T   A  M  K   Q  A  V   D  N  A   Y  A  R   G  V  V
694   GGACCATCA GGCTCAACA GCGATGAAA CAGGCGGTT GACAATGCA TATGCAAGA GGGGTTGTC

      V  V  A   A  A  G   N  S  G   S  S  G   N  T  N   T  I  G   Y  P  A
757   GTTGTGGCG GCTGCTGGG AACAGCGGA TCTTCAGGA AACACGAAT ACAATCGGC TATCCTGCG

      K  Y  D   S  V  I   A  V  G   A  V  D   P  N  S   N  R  A   S  F  S
820   AAATACGAC TCTGTCATC GCAGTTGGC GCGGTAGAC CCTAACAGC AACAGAGCT TCATTTTCC

      S  V  G   A  E  L   E  V  M   A  P  G   A  G  V   Y  S  T   Y  P  T
883   AGCGTCGGA GCAGAGCTT GAAGTCATG GCTCCTGGC GCAGGCGTG TACAGCACT TACCCAACC

      S  T  Y   A  T  L   N  G  T   S  M  A   S  P  H   V  A  G   A  A  A
946   AGCACTTAT GCAACATTG AACGGAACG TCAATGGCT TCTCCTCAT GTAGCGGGA GCAGCAGCT

      L  I  L   S  K  H   P  N  L   S  A  S   Q  V  R   N  R  L   S  S  T
1009  TTGATCTTG TCAAAACAT CCGAACCTT TCAGCTTCA CAAGTCCGC AACCGTCTC TCCAGTACG

      A  T  Y   L  G  S   S  F  Y   Y  G  K   G  L  I   N  V  E   A  A  A
1072  GCGACTTAT TTGGGAAGC TCCTTCTAC TATGGAAAA GGTCTGATC AATGTCGAA GCTGCCGCT

      Q  *
1135  CAATAA
```

DNA Sequence 5A36 *B.licheniformis* complete

FIG. 17A

```
     V  R  S  K  K  L  W  I  S  L  L  F  A  L  T  L  I  F  T  M  A
1    GTGAGAAGC AAAAAATTG TGGATCAGC TTGTTGTTT GCGTTAACG TTAATCTTT ACGATGGCG

     F  S  N  M  S  A  Q  A  A  G  K  S  S  T  E  K  K  Y  I  V  G
64   TTCAGCAAC ATGTCTGCG CAGGCTGCC GGAAAAAGC AGTACAGAA AAGAAATAC ATTGTCGGA

     F  K  Q  T  M  S  A  M  S  S  A  K  K  K  D  V  I  S  E  K  G
127  TTTAAACAG ACAATGAGT GCCATGAGT TCCGCCAAG AAAAAGGAT GTTATTTCT GAAAAAGGC

     G  K  V  Q  K  Q  F  K  Y  V  N  A  A  A  A  T  L  D  E  K  A
190  GGAAAGGTT CAAAAGCAA TTTAAGTAT GTTAACGCG GCCGCAGCA ACATTGGAT GAAAAAGCT

     V  K  E  L  K  K  D  P  S  V  A  Y  V  E  E  D  H  I  A  H  E
253  GTAAAAGAA TTGAAAAAA GATCCGAGC GTTGCATAT GTGGAAGAA GATCATATT GCACATGAA

     Y  A  Q  S  V  P  Y  G  I  S  Q  I  K  A  P  A  L  H  S  Q  G
316  TATGCGCAA TCTGTTCCT TATGGCATT TCTCAAATT AAAGCGCCG GCTCTTCAC TCTCAAGGC

     Y  T  G  S  N  V  K  V  A  V  I  D  S  G  I  D  S  S  H  P  D
379  TACACAGGC TCTAACGTA AAAGTAGCT GTTATCGAC AGCGGAATT GACTCTTCT CATCCTGAC

     L  N  V  R  G  G  A  S  F  V  P  S  E  T  N  P  Y  Q  D  G  S
442  TTAAACGTC AGAGGCGGA GCAAGCTTC GTACCTTCT GAAACAAAC CCATACCAG GACGGCAGT

     S  H  G  T  H  V  A  G  T  I  A  A  L  N  N  S  I  G  V  L  G
505  TCTCACGGT ACGCATGTA GCCGGTACG ATTGCCGCT CTTAATAAC TCAATCGGT GTTCTGGGC

     V  A  P  S  A  S  L  Y  A  V  K  V  L  D  S  T  G  S  G  Q  Y
568  GTAGCGCCA AGCGCATCA TTATATGCA GTAAAAGTG CTTGATTCA ACAGGAAGC GGCCAATAT

     S  W  I  I  N  G  I  E  W  A  I  S  N  N  M  D  V  I  N  M  S
631  AGCTGGATT ATTAACGGC ATTGAGTGG GCCATTTCC AACAATATG GATGTTATC AACATGAGC

     L  G  G  P  T  G  S  T  A  L  K  T  V  V  D  K  A  V  S  S  G
694  CTTGGCGGA CCTACTGGT TCTACAGCG CTGAAAACA GTCGTTGAC AAAGCCGTT TCCAGCGGT

     I  V  V  A  A  A  A  G  N  E  G  S  S  G  S  T  S  T  V  G  Y
757  ATCGTCGTT GCTGCCGCA GCCGGAAAC GAAGGTTCA TCCGGAAGC ACAAGCACA GTCGGCTAC

     P  A  K  Y  P  S  T  I  A  V  G  A  V  N  S  S  N  Q  R  A  S
820  CCTGCAAAA TATCCTTCT ACTATTGCA GTAGGTGCG GTAAACAGC AGCAACCAA AGAGCTTCA

     F  S  S  A  G  S  E  L  D  V  M  A  P  G  V  S  I  Q  S  T  L
883  TTCTCCAGC GCAGGTTCT GAGCTTGAT GTGATGGCT CCTGGCGTG TCCATCCAA AGCACACTT

     P  G  G  T  Y  G  A  Y  N  G  T  S  M  A  T  P  H  V  A  R  A
946  CCTGGAGGC ACTTACGGC GCTTATAAC GGAACGTCC ATGGCGACT CCTCACGTT GCCCGAGCA

     A  A  L  I  L  S  K  H  P  T  W  T  N  A  Q  V  R  D  R  L  E
1009 GCAGCGTTA ATTCTTTCT AAGCACCCG ACTTGGACA AACGCGCAA GTCCGTGAT CGTTTAGAA

     S  T  A  T  Y  L  G  N  S  F  Y  Y  G  K  G  L  I  N  V  Q  A
1072 AGCACTGCA ACATATCTT GGAAACTCT TTCTACTAT GGAAAAGGG TTAATCAAC GTACAAGCA

     A  A  Q  *
1135 GCTGCACAA TAA
```

DNA Sequence 3A1 *B.subtilis complete*

FIG 17. B

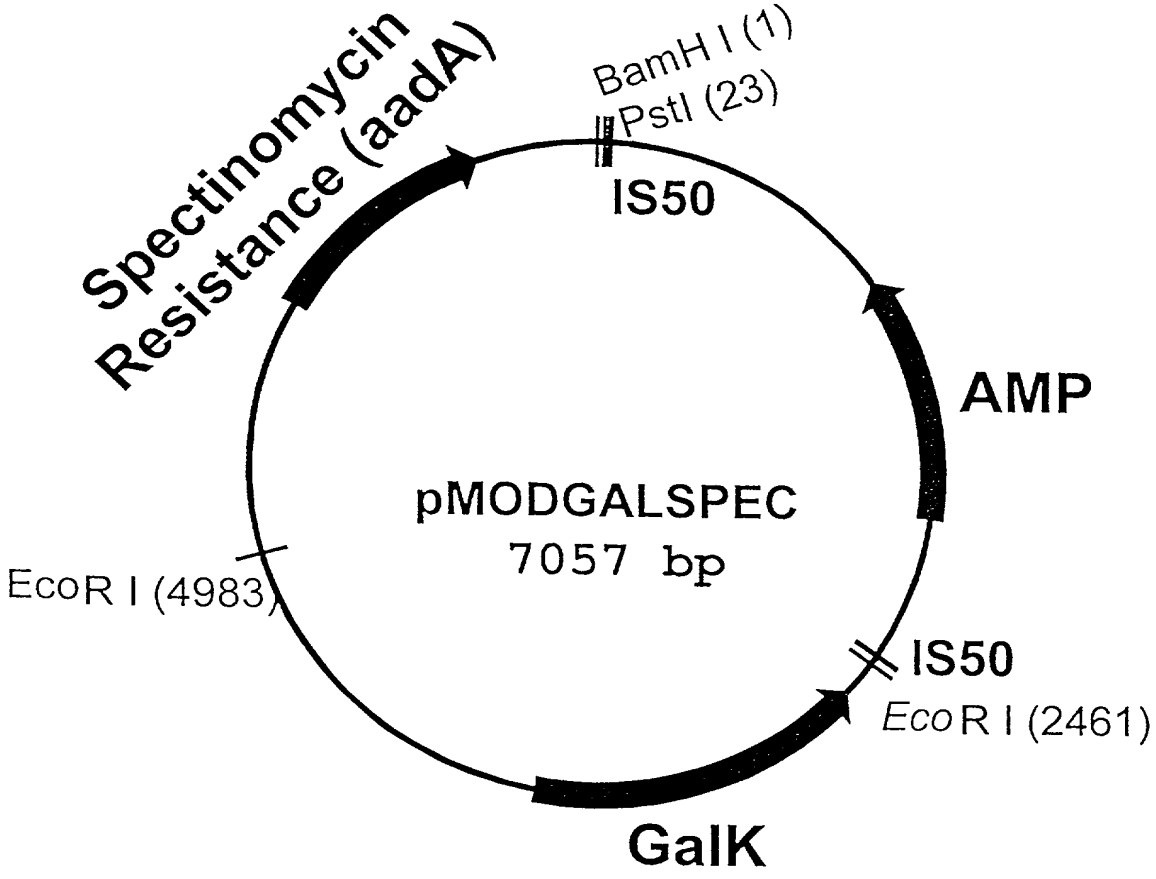

Spectinomycin Resistance (aadA)
BamH I (1)
PstI (23)
IS50
AMP
pMODGALSPEC
7057 bp
EcoR I (4983)
IS50
EcoR I (2461)
GalK

FIG. 18

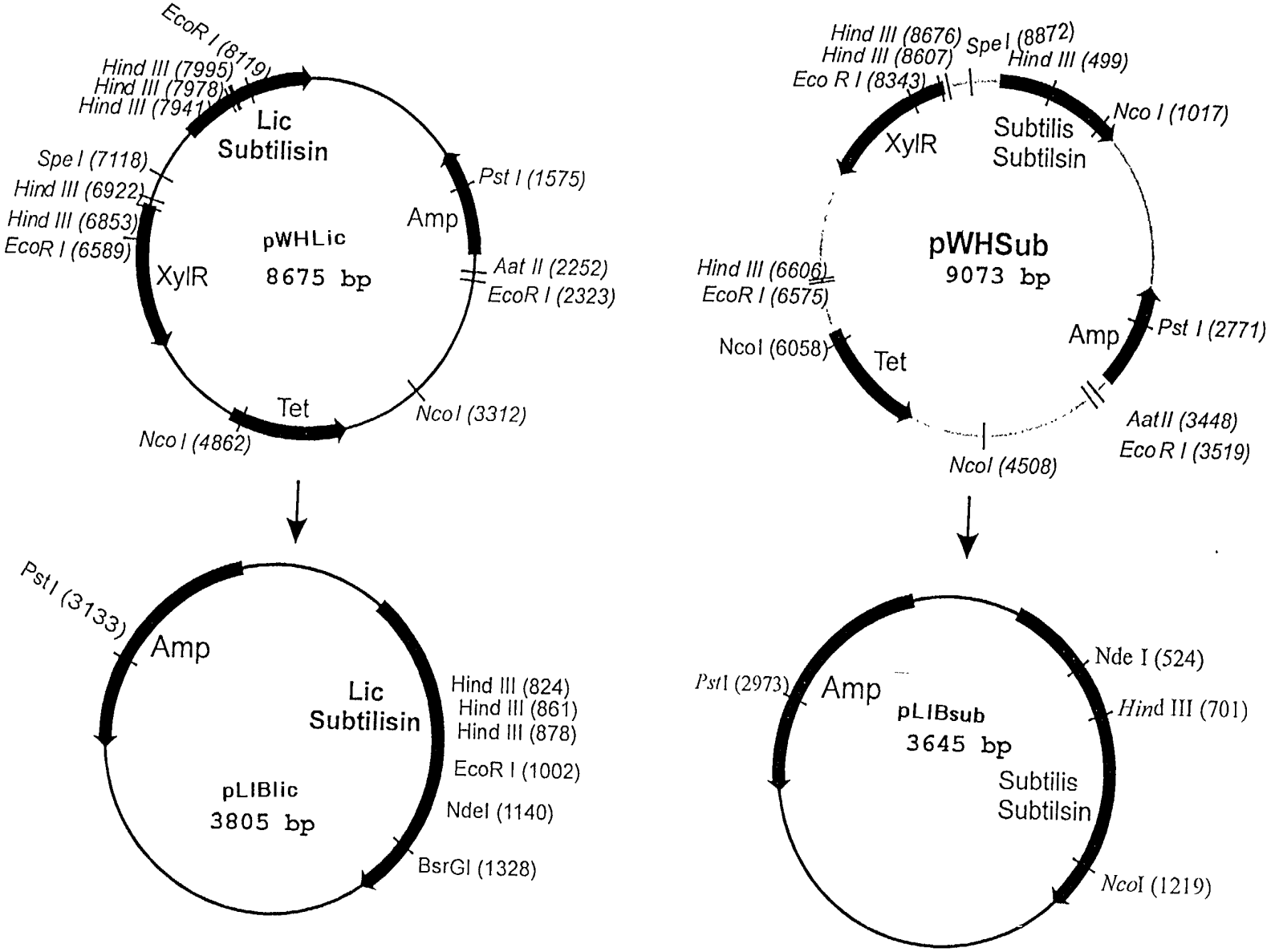

EcoR I (8779)
Hind III (7995)
Hind III (7978)
Hind III (7941)
Lic Subtilisin
Spe I (7118)
Hind III (6922)
Hind III (6853)
EcoR I (6589)
pWHLic
8675 bp
XylR
Amp
Pst I (1575)
Aat II (2252)
EcoR I (2323)
Tet
Nco I (4862)
Nco I (3312)
PstI (3133)
Amp
Lic Subtilisin
pLIBlic
3805 bp
Hind III (824)
Hind III (861)
Hind III (878)
EcoR I (1002)
NdeI (1140)
BsrGI (1328)
Hind III (8676)
Hind III (8607)
Spe I (8872)
Hind III (499)
Eco R I (8343)
XylR
Subtilis Subtilsin
Nco I (1017)
pWHSub
9073 bp
Hind III (6606)
EcoR I (6575)
NcoI (6058)
Tet
Amp
Pst I (2771)
AatII (3448)
EcoR I (3519)
NcoI (4508)
PstI (2973)
Amp
pLIBsub
3645 bp
Subtilis Subtilsin
Nde I (524)
Hind III (701)
NcoI (1219)

FIG. 19

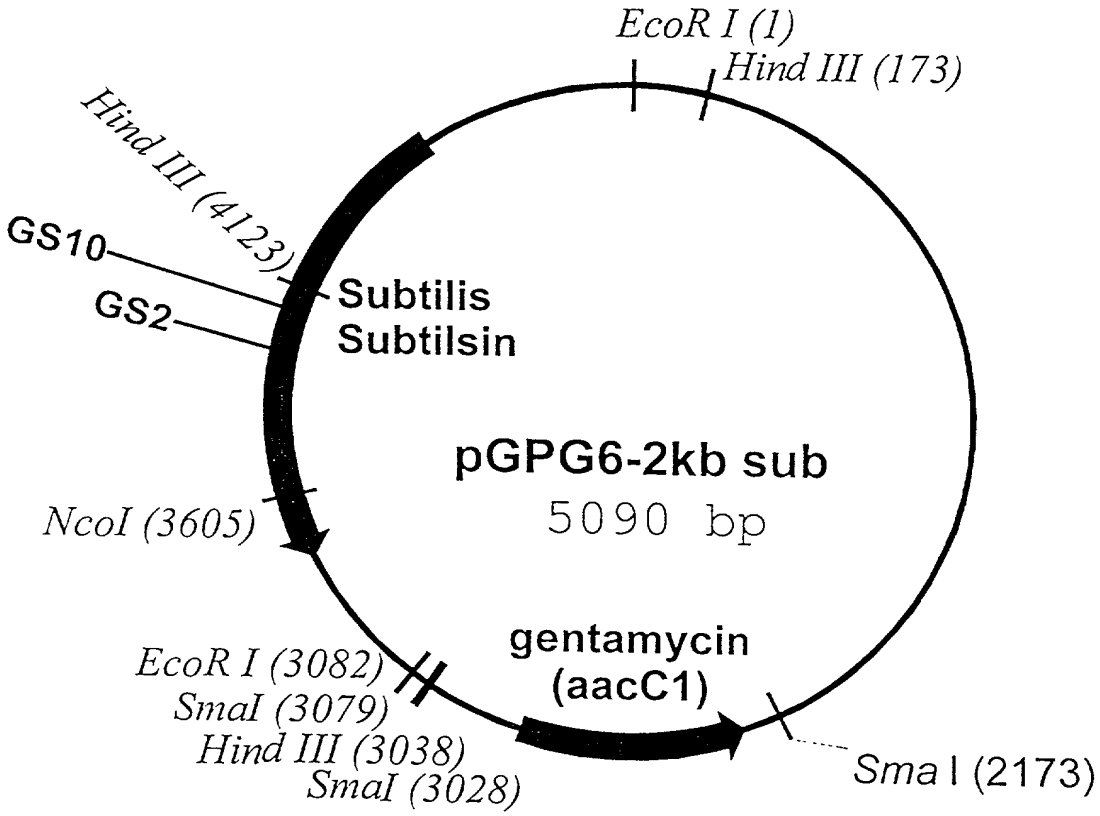
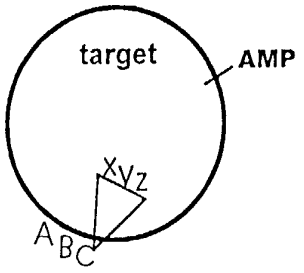
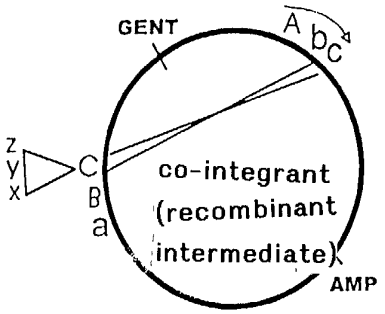

EcoR I (1)
Hind III (173)
Hind III (4123)
GS10
GS2
Subtilis
Subtilsin
pGPG6-2kb sub
5090 bp
NcoI (3605)
gentamycin
(aacC1)
EcoR I (3082)
SmaI (3079)
Hind III (3038)
SmaI (3028)
Sma I (2173)

FIG. 20

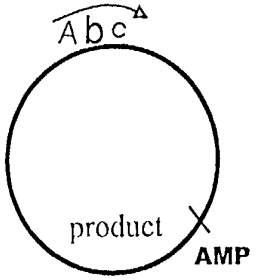
From the co-integrant
possible molecules include:



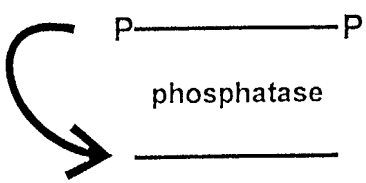
Enzyme digesting
uniquely in the insert (xyz)
will cut



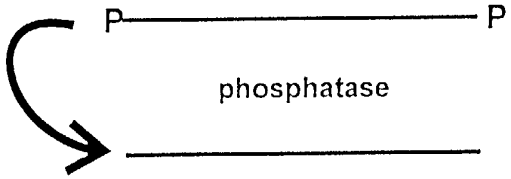
Enzyme digesting
uniquely in the insert (xyz)
will cut



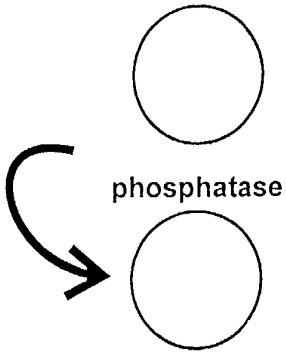
Enzyme digesting
uniquely in the insert (xyz)
doesnot cut



very few tranformants



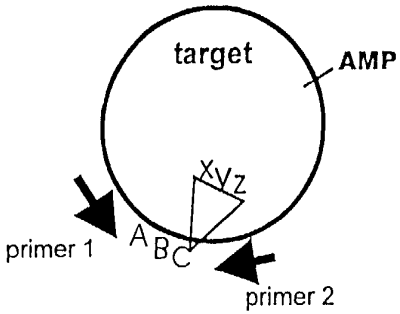
very few tranformants



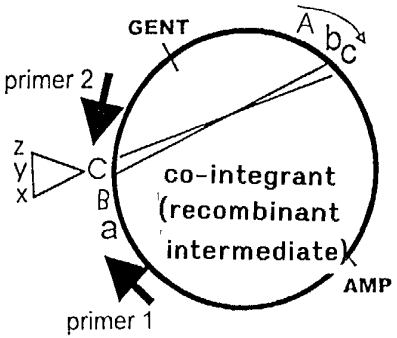
normal number of
tranformants

FIG. 21

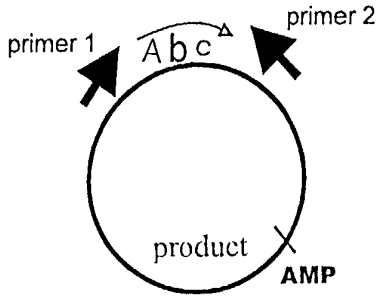
From the co-integrant
possible molecules include:



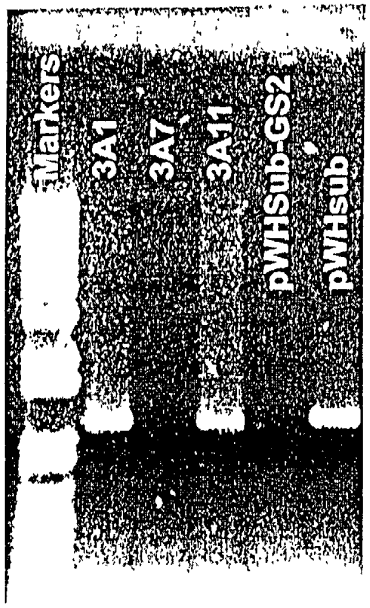
no product
(too long)



no product
(too long)



product



← PCR product

FIG. 22

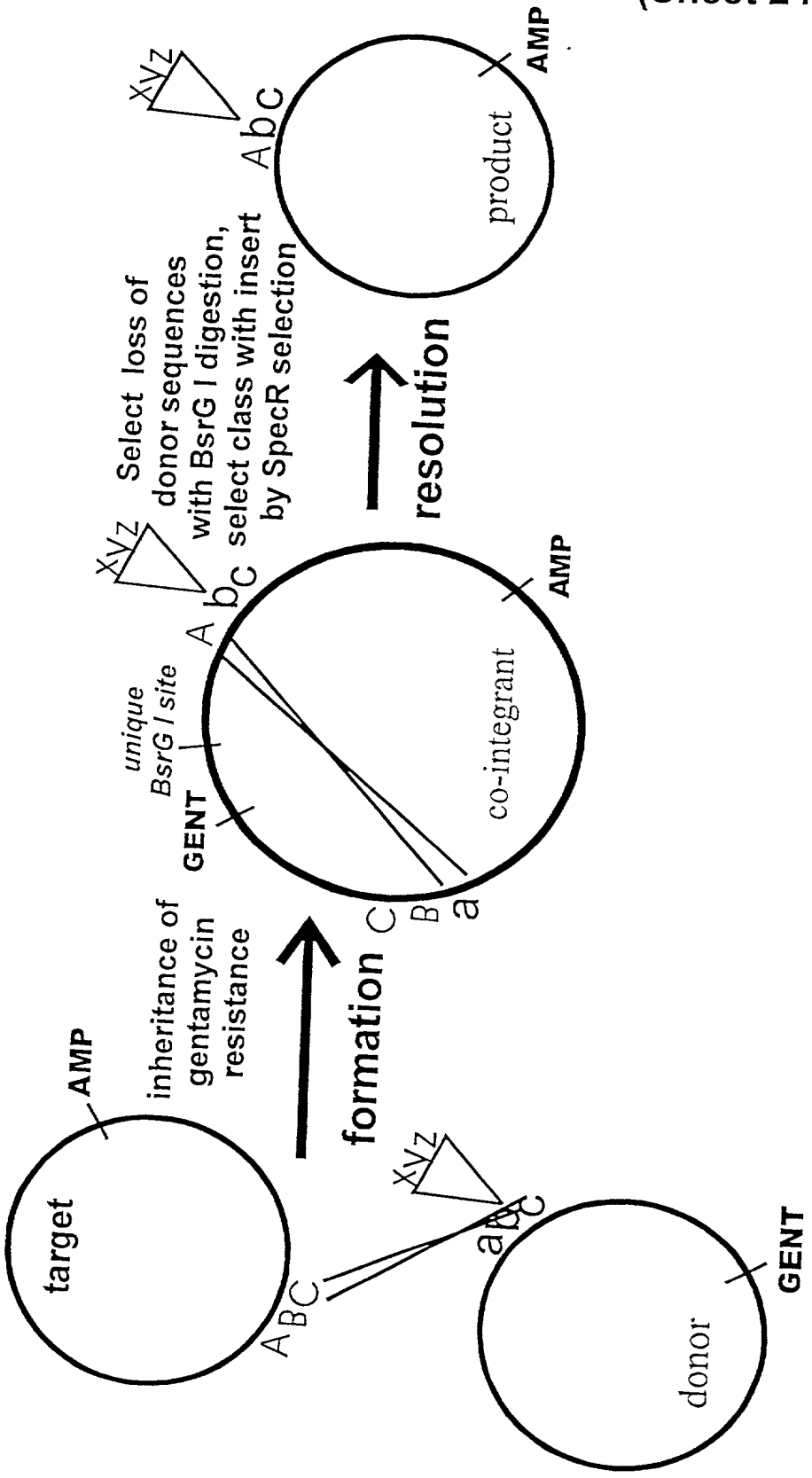

target
AMP
inheritance of gentamycin resistance
formation
donor
GENT
co-integrant
unique BsrG I site
GENT
AMP
Select loss of donor sequences with BsrG I digestion, select class with insert by SpecR selection
resolution
product
AMP

FIG. 23

Oligonucleotide Name	Oligonucleotide SEQ ID NO.	Sequence of oligonucleotide
MCS1F	SEQ ID NO:1	5'-AATTCGCGTTTAAACTTAATTAAGGTACCCATTTTTTGGCAGATCTAGACCAAAAAA TGGGGGCGGCCGCTCCCCGGGTGGCGCGCC-3'
MCS1R	SEQ ID NO:2	5'-AATTGGCGCGCCACCCGGGGAGCGGCCGCCCCCATTTTTTGGTCTAGATCTGCCAAA AAATGGGTACCTTAATTAAGTTTAAACGCG-3'
BglKF	SEQ ID NO:3	5'-GACTGCGAGATCATAGATATAGATTTCACTACGCGGCTGCTCAAACCTGG-3'
BglKR	SEQ ID NO:4	5'-CCAGGTTTGAGCAGCCGCGTAGTGAAATCTATATCTATGATCTCGCAGTC-3'
CC_UPPER	SEQ ID NO:5	5'-AATTTACCATGGAGCAATTGCATATGGTTTAAACAGCTCGAGTAGATCTTGCGGCC GCTTGGCTAGCGTCAGCTGGGTACCATGCAT -3'
CC_LOWER	SEQ ID NO:6	5'-CGCGTTATGCATGGTACCCAGCTGACGCTAGCCAAGCGGCCGCAAGATCTACTCGAG CTGTTTAAACCATATGCAATTGCTCCATGG-3'
internal primer - upper	SEQ ID NO:7	5'-CGCAA(T/A)C(T/C)GTTCCTTA(C/T)GG-3'
internal primer - upper	SEQ ID NO:8	5'-GCCAGGAGCCAT(C/G)AC(A/T)TCAA-3'

FIG. 24A

Oligonucleotide Name	Oligonucleotide SEQ ID NO.	Sequence of oligonucleotide
B. lichenformis Subtilisin Forward Primer	SEQ ID NO:9	5'-GGGGTACCGCGGTCTATTCATACTTTCG-3'
B. lichenformis Subtilisin Reverse Primer	SEQ ID NO:10	5'-GCAGATCTCATTTGTTAGAATATGTTATTGAGCGGC-3'
B. subtilis Subtilisin Forward Primer	SEQ ID NO:11	5'-AGCGAGATCTCTATTATTGTGCAGCTG-3'
B. subtilis Subtilisin Reverse Primer	SEQ ID NO:12	5'-GCGCGGTACCTGATAAAAGGAGAGGGTAAAGAG-3'
Galactokinase upper primer	SEQ ID NO:17	5'-GGAAGATCTAGAGGTTTTCACCGTCATCACCG-3'
Galactokinase lower primer	SEQ ID NO:18	5'-GGTAGATCTCTTTCGTCGTCTTCAAGAATTCCGC-3'

FIG. 24B